L5. Wasserstein distances

© Jan Obłój

$(\mathcal{X}, d)$ Polish. $p \geq 0$ with $d(x,y)^0 = 1_{x \neq y}$ by convention.

Let $T_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y)^p \, \pi(dx\,dy)$

Fix $x_0 \in \mathcal{X}$. Let $\mathcal{P}_p(\mathcal{X}) = \{\mu \in \mathcal{P}(\mathcal{X}) : \int d(x_0, x)^p \mu(dx) < \infty\}$.

Thm (Wasserstein distances) For all $p \geq 1$, $W_p := T_p^{1/p}$ is a metric on $\mathcal{P}_p(\mathcal{X})$.

For all $p \in [0,1)$, $W_p := T_p$ is a metric on $\mathcal{P}_p(\mathcal{X})$.

Rk If d is bdd, e.g. $\tilde{d} = d \wedge 1$, then $\mathcal{P}_p(\mathcal{X}) = \mathcal{P}(\mathcal{X})$.

Rk 2 $(\mathcal{X}, \|\cdot\|)$ Hilbert; $a \in \mathcal{X}$ then $W_2(\mu, \delta_a)^2 = \int_{\mathcal{X}} \|x - a\|^2 d\mu$

$\Rightarrow \int x \, d\mu =: m$ solves $\min_{a \in \mathcal{X}} W_2(\mu, \delta_a)$

& the cost is $\mathrm{Var}(\mu)$.

Proof (Case $p \in [0,1)$ follows from $p=1$ replacing d by topologically equivalent d^p.)

$W_p(\mu,\nu) = W_p(\nu,\mu)$, $W_p(\mu,\nu) \geq 0$ & finite on $\mathcal{P}_p(\mathcal{X})$.

$W_p(\mu,\mu) = 0$. Conversely, $W_p(\mu,\nu) = 0$ we take $\pi^* \in \Pi(\mu,\nu)$ then

$d(x,y) = 0$ $d\pi$-a.e. $\Rightarrow \pi(\{(x,x) : x \in \mathcal{X}\}) = 1$

$\Rightarrow \int \varphi(x) d\mu = \int_{\mathcal{X} \times \mathcal{Y}} \varphi(x) \pi(dx\,dy) = \int \varphi(y) \pi(dx\,dy) = \int \varphi(y) \nu(dy)$ $\forall \varphi \in C_b$

$\Rightarrow \mu = \nu$.

Finally we check the Δ-ineq. Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(\mathcal{X})$

Take optimal π_{12}^*, π_{23}^* & use glueing lemma to get $\pi \in \Pi(\mu_1, \mu_2, \mu_3)$ so $\pi_{13} \in \Pi(\mu_1, \mu_3)$.

$$W_p(\mu_1, \mu_3) \leq \left(\int_{X_1 \times X_3} d(x_1, x_3)^p \, d\pi_{13}(x_1, x_3) \right)^{1/p}$$

$$= \left(\int_{X_1 \times X_2 \times X_3} d(x_1, x_3)^p \, d\pi(x_1, x_2, x_3) \right)^{1/p}$$

$$\leq \left(\int \left(d(x_1, x_2) + d(x_2, x_3) \right)^p d\pi \right)^{1/p}$$

Minkowski

$$\leq \left(\int d(x_1, x_2)^p \, d\pi \right)^{1/p} + \left(\int d(x_2, x_3)^p \, d\pi \right)^{1/p}$$

$$= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$$

☺

Rk $p_1 \geq p_2 \geq 1 \implies W_{p_1} \geq W_{p_2}$ by Hölder.

Thm Let $p \geq 1$ & $(\mu_n) \subseteq \mathcal{P}_p(X)$, $\mu \in \mathcal{P}(X)$. TFAE

(i) $W_p(\mu_k, \mu) \to 0$ as $k \to \infty$

(ii) $\mu_k \to \mu$ & $\int d(x_0, x)^p \, d\mu_k \to \int d(x_0, x)^p \, d\mu < \infty$

(iii) $\forall \varphi$ Cont. $|\varphi(x)| \leq C(1 + d(x_0, x)^p)$ for some $x_0 \in X$, $C \in \mathbb{R}$

$$\int \varphi \, d\mu_k \to \int \varphi \, d\mu < \infty$$

Rk $\forall \varepsilon > 0 \ \exists C_\varepsilon > 0 \ \forall a, b \quad (a+b)^p \leq (1+\varepsilon) a^p + C_\varepsilon b^p$

$$\implies d(x_0, x)^p \leq (1+\varepsilon) d(x_0, y)^p + C_\varepsilon d(x, y)^p \quad \forall x, y, x_0$$

$\Rightarrow$ if $\eta \in \mathcal{P}_p(\mathcal{X})$ & $W_p(\eta, \mu) < \infty \Rightarrow \mu \in \mathcal{P}_p(\mathcal{X})$

since $\int d(x_0, x)^p \mu(dx) = \int d(x_0, x)^p \pi^*(dx\,dy) \leq (1+\varepsilon) \int d(x_0, y)^p \, d\pi^* + C_\varepsilon \underbrace{\int d(x,y)^p \, d\pi^*}_{= W_p(\mu,\eta)^p} < \infty.$

Rk "Wasserstein metrises weak conv."

Proof

(ii') $\Rightarrow$ (ii) ✓.

(ii) $\Rightarrow$ (ii') ? Take such $\varphi = \varphi^+ - \varphi^-$ & deal separately so assume $\varphi \geq 0$.

Let $\varphi_R = \varphi \wedge C(1+R^p)$

$$\int \varphi \, d\mu_k = \int \varphi_R \, d\mu_k + \int \underbrace{(\varphi - \varphi_R)}_{\leq C\, d(x_0,x)^p \mathbb{1}_{d(x_0,x) > R}} d\mu_k$$

$\downarrow$

$\int \varphi_R \, d\mu$

$$\left| \int \varphi \, d\mu_k - \int \varphi \, d\mu \right| \leq \left| \int \varphi_R \, d\mu_k - \int \varphi_R \, d\mu \right| + \underbrace{\int c\, d(x_0,x)^p \mathbb{1}_{d(x_0,x) > R} (d\mu + d\mu_k)}_{\leq \sup_k \text{ ``} \xrightarrow[R \to \infty]{} 0}$$

$k \to \infty$ $\downarrow$ by weak cv. 0

useful $\int d^p \wedge R \, d\mu_k \to \int d^p \wedge R \, d\mu$

& $\int d^p \, d\mu_k \to \int d^p \, d\mu$.

Remains to show (i) $\Leftrightarrow$ (ii).

$$\mu_k \to \mu \Rightarrow \int d(x_0,x)^p \, d\mu \overset{\text{MCT + weak conv}}{=} \lim_{R \to \infty} \lim_{k \to \infty} \int (d(x_0,x) \wedge R)^p \, d\mu_k$$
$$\leq \liminf \int d(x_0,x)^p \, d\mu_k$$

so (ii) is equivalent to $\mu_k \to \mu$ & $\limsup \int d(x_0,x)^p \, d\mu_k \leq \int d(x_0,x)^p \, d\mu(x)$ $(*)$.
$\mu \in \mathcal{P}_p(\mathcal{X})$

Using $\pi^* \in \Pi(\mu_k, \mu)$ on $d(x_0,x)^p \leq (1+\varepsilon)\, d(x_0,y)^p + C_\varepsilon\, d(x,y)^p$

$$\int d(x_0,x)^p \, d\mu_k \leq (1+\varepsilon) \int d(x_0,y)^p \, d\mu + C_\varepsilon W_p^p(\mu_k, \mu)$$

$\sup_u \quad - \ldots \quad - \ldots$

& letting $\varepsilon \searrow 0$ we get $(\ast)$.

So remains to show $W_p(\mu_u, \mu) \to 0 \Rightarrow \mu_u \to \mu$ & (ii) $\Rightarrow$ (i).

Taking $\tilde{d} = d \wedge 1$ one shows it is enough to establish these for a bounded d.

$\Rightarrow$ All W_p are equivalent so we work with $p = 1$ & $d \leq 1 \Rightarrow$

$$W_1(\mu, \nu) = \sup_{\substack{\varphi : \|\varphi\|_{Lip} \leq 1 \\ \|\varphi\|_\infty \leq 1}} \int \varphi \, d(\mu_u - \mu)$$

(i) $\Rightarrow \int \varphi \, d\mu_u \to \int \varphi \, d\mu \quad \forall \varphi$ Lipschitz (take $\frac{\varphi}{\|\varphi\|_{Lip}}$).

but then any $\varphi \in C^b$, $\exists$ unif bdd sequences of Lip functions a_n, b_n

$$\lim \nearrow a_n = \varphi = \lim \searrow b_n$$

$$\overline{\lim} \int \varphi \, d\mu_u \leq \lim \overline{\lim} \int b_n \, d\mu_u = \lim \int b_n \, d\mu_u = \int \varphi \, d\mu$$

& $\underline{\lim} \int \varphi \, d\mu_u \geq \int \varphi \, d\mu$ ✓. So (i) $\Rightarrow$ (ii).

Remains to argue the converse. Let $\mu_u \to \mu$.

We shift all 1-Lip φ so that $\varphi(x_0) = 0$.

We use Prokhorov $\Rightarrow \exists (K_n) \quad \sup_u \mu_u(K_n^c) \vee \mu(K_n^c) \leq 1/n$

AA $\Rightarrow \{\varphi 1_{K_n} : \varphi \in Lip_1(\mathcal{X}), \varphi(x_0) = 0\}$ is compact in $C_b(K_n)$

diagonal argument $\forall$ sequence (φ_u), $\exists \varphi_{u_m} \to \varphi_\infty$ unif. on each K_n,

& φ_∞ is bdd & Lip since all (φ_u) are unif bdd & unif 1-Lip.

Take φ_u s.t. $\sup \int \varphi \, d(\mu_u - \mu) \leq \int \varphi_u \, d(\mu_u - \mu) + 1/u$

φ_∞ is defined on $\bigcup K_n$. Extend this 1-Lip function to $\mathcal{X}$ via

$$\varphi_\infty(x) = \inf_{y \in \bigcup K_n} (\varphi_\infty(y) + d(x, y)) \quad \text{again 1-Lip.}$$

$$\int \varphi_u \, d(\mu_u - \mu) \leq \left| \int_{K_n} (\varphi_u - \varphi_\infty) \, d(\mu_u - \mu) \right| + \left| \int_{K_n^c} (\varphi_u - \varphi_\infty) \, d(\mu_u - \mu) \right|$$

$\leq 2 \cdot C \cdot 1/n \to 0$

$$\times \left[\int_X \varphi_\infty \, d(\mu_k - \mu) \right] \longrightarrow 0$$

$k \to \infty$: $\to 0$ by unif conv. ; $k \to \infty$: $\to 0$ by weak conv.

so $W_1(\mu_k, \mu) = \sup_{\varphi : \dots} \int \varphi \, d(\mu_k - \mu) \longrightarrow 0$

A somewhat different proof of the above uses the following result which will be useful to us later. (from now on $p \geq 1$)

Lemma Let (μ_k) be Cauchy in $(\mathcal{P}_p(X), W_p)$. Then (μ_k) is tight.

Thm (TV control)

$$W_p(\mu, \nu) \leq 2^{1/q} \left(\int d(x_0, x)^p \, d|\mu - \nu|(x) \right)^{1/p}, \quad \frac{1}{q} + \frac{1}{p} = 1$$

Thm $(\mathcal{P}_p(X), W_p)$ is Polish.

Proof · **Separability**: Let $Q = \left\{ \sum_{i=1}^{n} a_i \delta_{x_i} : n \in \mathbb{N}, a_i \in \mathbb{Q}, x_i \in D \right\}$

for $D \subseteq X$ countable dense.

For $\mu \in \mathcal{P}_p(X)$, $\varepsilon > 0$, take compact K : $\int_{X \setminus K} d(x_0, x)^p \, d\mu \leq \varepsilon^p$.

Cover K with $B(x_k, \varepsilon/2)$, $k \leq N$, $x_k \in D$.

$$B'_k = B(x_k, \varepsilon) \setminus \bigcup_{j<k} B(x_j, \varepsilon) \quad \text{disjoint cover.}$$

Let $f : X \to X$ $\quad f|_{B'_k \cap K} = x_k$ & $f|_{X \setminus K} = x_0$

so $\forall x \in K \quad d(x, f(x)) \le \varepsilon$ &

$$\int d(x, f(x))^p \, d\mu \le \varepsilon^p \int_K d\mu + \int_{X \setminus K} d(x, x_0)^p \, d\mu \le 2\varepsilon^p.$$

so $W_p(\mu, f\#\mu) \le 2\varepsilon^2$ & $f_\#\mu$ is a finite sum of Diracs.

Now $W_p\left(\sum_i a_i \delta_{x_i}, \sum_i b_i \delta_{x_i}\right) \le 2^{1/p} \max_{k,l} d(x_k, x_l) \sum_{j \le N} |a_j - b_j|^{1/p}$

& we can approximate $f\#\mu$ with $\mu_\varepsilon \in Q$.

- <u>Completeness</u>

Take a Cauchy sequence $\Rightarrow$ tight $\Rightarrow \mu_{k_n} \longrightarrow \mu$.

$$\int d(x_0, x)^p \, d\mu \le \underline{\lim} \int d(x_0, x)^p \, d\mu_k < \infty$$

by Cauchy

W_p is the value of OT for a cont cost $\Rightarrow$ is l.s.c. so

$$W_p(\mu, \mu_l) \le \underline{\lim}_k W_p(\mu_k, \mu_l)$$

$$\Rightarrow \limsup_l W_p(\mu, \mu_l) \le \overline{\lim}_{k,l \to \infty} W_p(\mu_k, \mu_l) = 0.$$

So $\mu_l \longrightarrow \mu$ in W_p.

Cauchy with a cv subsequence is converging.

$\blacksquare$